引用格式: 梅高兴，肖寻，陈仕语，等． 中文版工作记忆容量测量任务— “自动化运算广度任务”的信效度检验［J］ ． 贵州师范大学学报(自然科学版) ，2021，39(4) : 98-103． ［MEI G X，XIAO X，CHEN S Y，et al． Reliability and validity of a Chinese version of the automated version of the operation span task［J］ ． Journal of Guizhou Normal University (Natural Sciences) ，2021，39(4) : 98-103］

中文版工作记忆容量测量任务— “自动化运算广度任务”的信效度检验

梅高兴，肖寻，陈仕语，苟丽娜，李席英

（贵州师范大学心理学院，贵阳，中国，550025）

摘要:采用基于 MATLAB 软件环境开发的简体中文版的“自动化运算广度任务”和瑞文推理测验对 102 名大学生进行施测，其中 86 名被试在约一周后进行重测。结果显示：简体中文版“自动化运算广度任务”的 Cronbach α 内部一致性信度系数为 0. 670，重测信度为 0.666；采用瑞文推理测验作为效度效标，将瑞文推理测验得分与总广度得分的皮尔逊相关分析显示，二者在初测（$r = 0.33, P < 0.01$）和重测（$r = 0.39, P < 0.001$）均呈显著正相关。结论：该简体中文版的“自动化运算广度任务”具有较好的信效度，操作简便，研究者可方便地测量个体的工作记忆容量。该任务可以在 https://www.labxing.com/lab/695/data 网站免费下载。

关 键 词: 工作记忆; 工作记忆容量; 自动化运算广度任务; 信度; 效度

Reliability and validity of a Chinese version of the automated version of the operation span task

Abstract: Using the simplified Chinese version of Automated Operation Span Task developed by the current study and the Raven Progressive Matrices, we tested 102 college students. Eighty-six participants were re-tested after approximately one-week. Results: The Cronbach’s α was 0.670, and the test-retest reliability was 0.666. Using the Raven Progressive Matrices as the criterion, we found that Pearson correlation analysis showed significantly positively correlated whenever in the initial test ($r = 0.33$, $P <0.01$) or the re-tested ($r = 0.39$, $P < 0.001$). Conclusion: The simplified Chinese version of the automated version of the operation span task had good reliability and validity, and was easy to operate for users. The task can be freely downloaded from https://www.labxing.com/lab/695/data.

Key words: working memory; working memory capacity; automated version of the operation span task; reliability; validity

0 引言

工作记忆是指个体对信息进行暂时储存和加工的容量有限的系统[1-2]。个体的工作记忆容量如智力和人格一样存在着广泛的个体差异。有大量研究表明，个体的工作记忆容量与其他高级认知能力息息相关，如流体智力[3]、推理能力[4]、阅读理解能力[5]等。工作记忆容量作为测量学指标也被广泛用于临床心理学、社会心理学、语言心理学、教育心理学、发展心理学等研究领域。例如，发展心理学的研究表明，工作记忆容量的改变是儿童认知发展和老年人认知老化的主要驱动力之一[6-7]；临床心理学的研究表明，工作记忆能力的减弱是早期阿兹海默症（Alzheimer's disease）的征兆之一[8]。所以，开发普遍适用的测量工作记忆容量的任务可以大力推进工作记忆相关领域的研究。

工作记忆广度任务是测量个体工作记忆容量的主要任务。工作记忆广度任务是在对短时记忆容量测量研究的基础上逐步发展而来的。早期对短时记忆容量的测量主要是采用短时记忆广度任务[9]。在这些短时记忆广度任务中，实验者首先给被试依次呈现一些要记忆的项目（如数字、字母、词语等），然后要求被试按顺序回忆出这些项目。这样的短时记忆广度任务主要是测量个体对暂时激活信息的储存能力，并没有涉及到对信息的加工。研究表明，这种短时记忆广度任务并不能够很好预测个体其他方面的高级认知能力，如阅读理解能力[10]。在短时记忆测量的基础上发展起来的工作记忆广度任务的设计是基于 Baddeley 和 Hitch 关于工作记忆理论的核心观点[2]：工作记忆包括信息储存和信息加工两部分。自 Daneman 和 Carpenter 按照 Baddeley 和 Hitch 的工作记忆理论设计出第一个工作记忆广度任务以来[2,11]，不同的研究者设计出了不同的工作记忆广度任务，主要包括：阅读广度任务 [11]、计数广度任务[12]、运算广度任务[13]等。这些工作记忆广度任务一般都包含两个部分：需要记忆的项目和需要加工的项目。不同任务的不同点主要在于要记忆的项目和需加工的项目在具体形式上有所不同。记忆项目的具体形式包括数字、字母、单词、形状和空间位置等。在实验时一般都要求被试按正确的顺序回忆。加工项目的具体形式包括出声阅读句子、计算数学算式、计数圆点数等。

工作记忆广度任务作为一种筛选被试（如挑选高工作记忆广度者或低工作记忆广度者）的工具已广泛应用到心理学的各个研究领域。但是，早期的工作记忆广度测量任务不能够满足大量挑选被试的要求，不便于进行团体测量，耗时耗力[13]。为克服这一缺点，研究者在“运算广度任务”的基础上开发了一种基于在心理学研究领域广泛使用的软件 E-prime 的任务程序—“自动化运算广度任务”（Automated version of operation span task）[14]。该程

序任务符合测量学的标准，具有较高的信度和效度，易于实验者施测，实验过程中极少需要实验者指导被试，被试通过鼠标点选就可以完成任务，有利于进行团体测量。基于这些优点，自该测量任务开发至今，作为工作记忆容量测量的工具，已广泛应用于心理学各研究领域[15-18, 26-28]。

为使中文研究者更方便地使用“自动化运算广度任务”测量工作记忆容量，此研究的目标是开发出基于一个更广泛使用的编程软件-MATLAB-的简体中文版“自动化运算广度任务”程序，介绍该任务程序的使用方法，并报告其试用情况。

1 方法

1.1 被试

采用方便取样随机招募被试 102 名(M = 20.3 岁, SD =1.5, 年龄范围为 18～24 岁，男生 27 人，女生 75 人)。参照 Unsworth 等开发的“自动化运算广度任务”成绩计算标准[14]，排除 14 名在完成自动化运算广度任务中数学算式任务的计算正确率低于 85%的被试，约占总人数的 14% （14/102）。这一结果与 Unsworth 等研究中未达到正确率标准的人数比例（15%）相似[14]。这部分被试未允许参与重测。有 2 名被试在第一次测试中正确率达标，但未参加第二次测试。剩余的 86 名被试参加“自动化运算广度任务”的重测，2 次测试平均间隔时间约为 8 天（间隔范围 5～11 天）。参与重测的被试中有 6 名被试的数学算式计算正确率低于 85%。初测和重测中计算正确率未达到 85%标准的被试，其工作记忆容量测试数据未被纳入统计分析。所有被试视力或矫正视力正常，并在完成测试后获得一定报酬。加上被试阅读指导语和练习的时间，被试完成“自动化运算广度任务”的平均时间约为 20 min。

1.2 测试任务和实验程序

“自动化运算广度任务”采用自主开发的简体中文版程序，该程序采用 MATLAB 软件和 PsychToolbox-3 工具箱编写[19-20]。在这个任务程序中被试需先依次完成 3 个练习。练习一是字母回忆任务。首先在屏幕中央依次呈现 2～3 个字母（从 F, H, J, K, L, N, P, Q, R, S, T 和 Y 这 12 个字母中随机选择），每个字母呈现时间为 0.8 秒。要求被试记住呈现的字母及其顺序。然后呈现一个 4 × 3 的字母矩阵(之前可能出现的所有字母，见图 1)，被试通过鼠标点选字母对应的空白框对记忆的字母按顺序进行回忆（被试点选空白框后，对应的字母会出现在屏幕的中下部）；无法回忆的字母可点选“空白”框代替。然后任务程序将对被试在这一组记忆字母的成绩给出反馈。被试在练习一完成 4 组字母记忆任务。

练习二是简单数学算式计算任务。首先在屏幕中央呈现 1 个简单的数学算式（如（5 × 2）- 6=？），被试需尽快计算结果并点击鼠标继续。接着屏幕上出现 1 个答案以及“正确”

和“错误”字样，被试通过鼠标判断正误。被试做出反应后任务程序将给予正误反馈。练习二需完成 15 个数学算式计算任务。每名被试在练习二中对数学算式平均反应时的 2.5 倍将作为在正式测试中数学算式呈现的最长时间。

练习三是将字母回忆任务与简单数学算式计算任务混合，正式测试的流程（见图 1）与练习三相同。屏幕首先呈现一个数学算式，被试计算出答案后点击鼠标到下一屏幕（当计算超时，程序也会自动跳到下一屏幕，算式呈现的最长时间来自于练习二）。在这一屏幕，一个数字被呈现（该数字既可能是正确答案也可能是错误答案），被试用鼠标点选屏幕上的“正确”或“错误”标签进行判断。然后 1 个字母呈现 0.8 秒，要求被试记忆。被试完成 3～7 次这样的双任务（数学计算任务和字母回忆任务）后，出现字母矩阵让被试回忆记忆的字母。然后任务程序将反馈被试在这一组记忆字母和计算数学算式的成绩，以及到目前为止所有数学算式计算的正确率（呈现在屏幕的右上角）。然后，下一组工作记忆容量测试任务开始。正式测试中被试将完成 15 组这样的测试。也就是，被试每组需记忆 3～7 个字母，共 5 种记忆广度，每种重复 3 次。每组中记忆广度的大小从 5 种记忆广度中随机选择 1 种。

正式测试共包含 75 个数学算式和 75 个字母，整个测试只需点击鼠标便可完成。

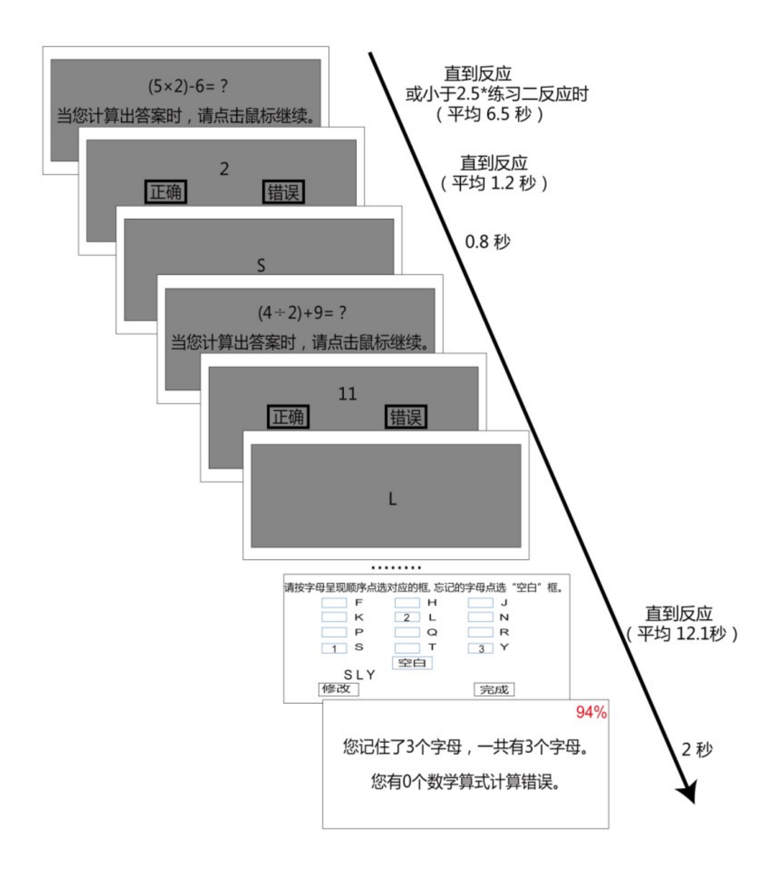


图 1 一组工作记忆容量测试流程图

Fig.1 A trial procedure of a set of working memory capacity test

为考察中文简体版“自动化运算广度任务”的效度，被试也需要完成瑞文推理测验[21]。研究使用的瑞文测验共计 36 个项目，分为 3 部分，每部分包含 12 个项目，每个项目都会缺失右下角的一张图案，被试通过观察该项目中的其它图案，寻找规律并从给出的 6～8 个选项中选出缺失图片。每部分的完成时间为 5 分钟。瑞文推理测验也采用 MATLAB 软件和 PsychToolbox-3 工具箱进行呈现。被试先完成“自动化运算广度任务”，然后再完成瑞文推理测验。

1.3“自动化运算广度任务”计分方法

根据以往的研究 [14,22-23]，在该任务中以总广度得分作为反映个体工作记忆容量大小的指标较好，因为绝对广度得分舍弃了部分试次信息，可能会导致无法区分被试表现。总广度得分是指被试在测试中回忆的字母和其位置顺序都正确的项目总和。例如，在一组被试需完成 3 次数学算式和记忆字母的测试中（如图 1 的样例），如果被试能够完全按顺序正确回忆出 3 个字母（如图 1 示例中的 SLY）, 则记 3 分；如果被试只回忆出 2 个字母且位置顺序正确（如 SL-,或-LY,或 S-Y）, 则记 2 分；如果回忆的字母错误或字母正确但位置顺序不正确（如 YSL）, 则均记 0 分。所以，在“自动化运算广度任务”下工作记忆容量的得分范围是 0～75 分。另外，任务程序还给出完全广度得分和错误总数（含计算超时错误数和计算错误数）。完全广度得分是指只有被试在某一组上所有字母和位置都回忆正确才计分。例如，在图 1 的样例中，只有被试完全按顺序正确回忆出所有的 3 个字母才记 3 分；其它情况均记 0 分。计算超时错误数是指在整个测试中被试计算数学算式时超出数学算式呈现时间的数量；计算错误数是指被试计算数学算式时计算错误的数量。错误总数是超时错误数和计算错误数的总和。

2 结果与讨论

2.1 描述统计

表 1 显示了被试在初测和重测中在“自动化运算广度任务”上的总广度得分、绝对广度得分、超时错误数和计算错误数的平均值、中位数、标准差、偏度、峰度、下四分位数和上四分位数。这些结果与 Unsworth 等[14]在开发该程序时的测量结果相似。如在 Unsworth 等人的初测测量结果中[14]，被试平均总广度得分和绝对广度得分分别为 55.25 和 39.16；目前研究中被试平均总广度得分和绝对广度得分分别为 55.91 和 38.16。 将数学算式计算正确率与工作

记忆的总广度得分进行皮尔逊相关检验，二者相关不显著（$r = 0.19, P > 0.05$），表明被试在完成数学算式和字母记忆时不存在权衡现象。

表 1 “自动化运算广度任务”各指标初测和重测的描述统计结果

Tab.1 Descriptive statistical results of all indexes of the automated version of the operation span for the initial and retest test

指标	平均值	中位数	标准差	偏度	峰度	下四分位数	上四分位数
总广度得分							
初测	55.91	57.00	10.12	-.98	3.45	50.00	63.00
重测	62.04	62.50	7.39	-.85	1.95	58.00	67.00
绝对广度得分							
初测	38.16	37.50	15.24	.27	.12	27.00	49.00
重测	46.88	48.00	13.12	.09	-.35	37.00	55.75
超时错误数							
初测	2.06	2.00	1.61	.82	.68	1.00	3.00
重测	2.44	2.00	2.17	.65	-.71	1.00	4.00
计算错误数							
初测	3.32	3.00	2.20	.44	-.56	2.00	5.00
重测	2.63	2.00	1.86	.75	.52	1.00	4.00

注:总样本 $N = 88$ 人;重测样本 $N = 80$ 人。因为重测中有 6 名被试数学算式正确率未达到 85%的正确率, 所以这些被试的数据为纳入重测相关指标的计算。

2.2 信度分析

对简体中文版“自动化运算广度任务”进行信度分析，包括内部一致性信度和重测信度分析。因为同一种工作记忆容量（3～7 个字母，5 种记忆广度）要随机重复测量 3 次（见方法部分），参照 Unsworth 等人的方法[14]，分别将每名被试在初测中完成第一次、第二次和第三次各种记忆广度的总广度得分合并为 1 个分数（也即每名被试有 3 个分数），然后用所有被试的 3 个分数计算出内部一致性系数（Cronbach α 系数）为 0.670。如果以绝对广度得分作为指标，内部一致性系数为 0.74。 另外，以初测和重测的总广度得分为指标，采用皮尔逊相关计算“自动化运算广度任务”的重测信度，结果显示相关显著（$r = 0.666, P < 0.001$）。 综上，内部一致性信度和重测信度的分析表明，基于 MATLAB 的简体中文版“自动化运算广度任务”具有良好的信度。

2.3 效度分析

参照以往的研究[3,14,23-25]，采用瑞文推理测验作为效标来考察中文版“自动化运算广度任务”的效度。被试的瑞文推理测验得分（$M = 13.07$, $SD = 4.38$）与总广度得分的皮尔逊相关分析显示，二者在初测（$r = 0.33$, $P < 0.01$）和重测（$r = 0.39$, $P < 0.001$）均呈显著正相关，被试瑞文推理测验得分越高，其工作记忆容量得分就倾向于越高。这表明中文版的“自动化运算广度任务”具有较好的效度。

3 讨论

Unsworth 等人开发的测量工作记忆容量的“自动化运算广度任务”是具有较高信、效度的测量工具[14]，应用广泛，但缺乏中文环境下的使用报告。为使国内研究者更方便的测量个体的工作记忆容量，本研究开发的基于 MATLAB 软件环境的简体中文版的“自动化运算广度任务”测验，具有较好的信度和效度，容易进行团体施测，测试个体只需通过鼠标点击即可完成整个测试任务。

本研究结果表明，中文版“自动化运算广度任务”的内部一致性系数为 0.67（以总广度得分为计分方式时）和 0.74（以绝对广度得分为计分方式时），说明该任务具有较好的内部一致性信度。在 Unsworth 等人开发该测量的研究中，该内部性一致性系数为 0.78，我们测量计算出的部性一致性系数略低于该值。另外，尽管本研究结果的重测信度（0.67）低于 Unsworth 等人开发该测量时的重测信度（0.83），但该重测信度在可接受的范围内。目前研究在这些指标上低于 Unsworth 等人开发该测量时所得到的这些指标，其原因可能是目前研究中使用的样本量偏少导致的。但是，总体上目前研究的信度指标基本满足心理测量学的要求。

采用瑞文推理测验作为效标，将总广度得分与瑞文推理测验得分进行相关分析，实验结果显示与瑞文推理测验呈显著正相关，这表明新开发的“自动化运算广度任务”具有较好的效标效度，这与前人的研究一致[14,26-27,29]。

研究开发的测量工作记忆容量的自动化运算广度任务（中文版）是基于 MATLAB R2013a 版本和 Windows 7 操作系统编写和测试的。在 Windows 10 操作系统进行了测试，可正常使用。运行该任务程序除安装 MATLAB 软件外，需添加 PsychToolbox-3 工具箱，该工具箱可在网站上免费下载：http://psychtoolbox.org/。该网站上有详细介绍如何添加 PsychToolbox-3 工具箱的流程供用户参考。“自动化运算广度任务”（中文版）可以在以下网站免费下载： https://www.labxing.com/lab/695/data。解压后运行“AospanChinese.m”脚本文件即可开始进行工作记忆容量的测试。测试完毕后，屏幕将显示主要测试结果（如总广度

得分），所有数据以 m 文件的形式保存在“Data”文件中。程序也将直接在命令窗口显示出相关测试结果；同时也会将主要测试结果单独生成 txt 文本资料存储在 Data 文件夹中。不熟悉 MATLAB 的用户可以直接复制 txt 文本资料中的主要结果到其他数据分析软件（如 SPSS）中进行进一步分析；熟悉 MATLAB 的用户可以撰写分析程序在储存的 m 文件中提取相关数据。程序运行中用户若想终止可按 Esc 键。

4 结论

研究开发的基于 MATLAB 软件环境的测量工作记忆容量的中文版“自动化运算广度任务”试用报告显示，该中文版任务具有较好的信度和效度，可以供各领域相关研究者方便地测量个体的工作记忆容量。在整个测量中被试只需要通过点击鼠标就可以完成。

参考文献：

[1]BADDELEY A. Working memory: Looking back and looking forward [J]. Nature Reviews Neuroscience, 2003, 4(10): 829-839.

[2]BADDELEY A D, HITCH G J. Working memory [M] New York: Academic Press, 1974: 47-89.

[3]KANE M J, HAMBRICK D Z, CONWAY A R. Working memory capacity and fluid intelligence are strongly related constructs: Comment on ackerman, beier, and boyle(2005) [J]. Psychological Bulletin, 2005, 131(1): 66-71.

[4]SÜß H-M, OBERAUER K, WITTMANN W W, et al. Working-memory capacity explains reasoning ability and a little bit more [J]. Intelligence, 2002, 30(3): 261-288.

[5]MCVAY J C, KANE M J. Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention [J]. Journal of Experimental Psychology: General, 2012, 141(2): 302-320.

[6]HERTZOG C, DIXON R A, HULTSCH D F, et al. Latent Change models of adult cognition: Are changes in processing speed and working memory associated with changes in episodic memory? [J].Psychology and Aging, 2003, 18(4): 755-769.

[7]KAIL R. Longitudinal evidence that increases in processing speed and working memory enhance children’s reasoning [J]. Psychological Science, 2007, 18(4): 312-313.

[8]DELLA S S, LOGIE R H. Theoretical and practical implications of uual-task performance in alzheimer’s disease [J]. Brain, 2001, 124(8): 1479-1481.

[9]DEMPSTER F N. Memory span: Sources of individual and developmental differences [J]. Psychological Bulletin, 1981, 89(1): 63-100.

[10]PERFETTI C A, LESGOLD A M. Discourse comprehension and sources of individual differences [M] Hillsdale, NJ: Erlbaum, 1977: 141-183.

[11]DANEAN M, CARPENTER P A. Individual differences in working memory and reading [J]. Journal of Verbal Learning and Verbal Behavior, 1980, 19(4): 450-466.

[12]CASE R, KURLAND M D, Goldberg J. Operational efficiency and the growth of short-term memory span [J]. Journal of Experimental Child Psychology, 1982, 33(3): 386-404.
[13]TURNER M L, ENGLE R W. Is working memory capacity task dependent? [J]. Journal of Memory & Language, 1989, 28(2): 127-154.
[14]UNSWORTH N, HEITZ R P, SCHROCK J C, et al. An automated version of the operation span task [J]. Behavior Research Methods, 2005, 37(3): 498-505.
[15]BENGSON J J, MANGUN G R. Spatial attention and feature-based attention are differentially sensitive to individual working memory capacity and perceptual load [J]. Visual Cognition, 2018, 26(7): 1-7.
[16]GAREAU A, CHAMANDY M, KLJAJIC K, et al. The detrimental effect of academic procrastination on subsequent grades: The mediating role of coping over and above past achievement and working memory capacity [J]. Anxiety, Stress & Coping, 2018, 32(2): 141-154.
[17]HATZ L E, MCCARTY K N, BARTHOLOW B D, et al. Explicit attitudes,working memory capacity and driving after drinking [J]. Alcoholism,Clinical and Experimental Rresearch, 2018, 42 (10): 2047-2053.
[18]KORHONEN J, NYROOS M, JONSSON B, et al. Additive and multiplicative effects of working memory and test anxiety on mathematics performance in grade 3 students [J]. Educational Psychology, 2017, 38(5): 572-595.
[19]BRAINAR D, DAVID H. The psychophysics toolbox [J]. Spatial Vision, 1997, 10(4): 433-436.
[20]PELL I, DENIS G. The video toolbox software for visual psychophysics: Transforming numbers into movies [J]. Spatial Vision, 1997, 10(4): 437-442.
[21]RAVEN J C, RAVEN J E, COURT J H. Raven manual section 4: Advanced progressive matrices [M]. Oxford:Oxford Psychologists Press, 1998.
[22]CONWAY A R A, KANE M J, BUNTING M F, et al. Working memory span tasks: A methodological review and user's guide [J]. Psychonomic Bulletin and Review, 2005, 12(5): 769-786.
[23]REDICK T S, BROADWAY J M, MEIER M E,et al. Measuring working memory capacity with automated complex span tasks [J]. European Journal of Psychological Assessment, 2012, 28(3): 164-171.
[24]GONTHIOR C, THOMASSIN N. Strategy use fully mediates the relationship between working memory capacity and performance on raven's matrices [J]. Journal of Experimental Psychology General, 2015, 144(5): 916-924.
[25]OBERAUER K, SCHULZE R, WILHELM O, et al. Working memory and intelligence their correlation and their relation: Comment on ackerman, beier, and boyle(2005) [J]. Psychological Bulletin, 2005, 131(1): 61-65.
[26]LEWANDOWSKY S, OBERAUER K, YANG LX, ECKER UK. A working memory test battery for MATLAB [J]. Behav Res Methods. 2010, 2(2): 571-85.
[27]ĐOKIĆ, RATKO, M KOSO-DRLJEVIĆ, ĐAPO, NERMIN, et al. Working memory span tasks: Group administration and omitting accuracy criterion do not change metric characteristics [J]. PLoS ONE. 2018, 13(10).

[28]UNSWORTH, N., FUKUDA, K., AWH, E., & VOGEL, E. K. Working memory and fluid intelligence: capacity, attention control, and secondary memory retrieval [J]. Cognitive Psychology, 2014, 71: 1–26.

[29]FOSTER JL, SHIPSTEAD Z, HARRISON TL, et al. Shortened complex span tasks can reliably measure working memory capacity [J]. Mem Cognit. 2015, 43(2):226-36.